

Semantic structures within natural language and their cognitive functions

Limin Zhang

Limin_IR@outlook.com

October 9, 2024

Abstract

Natural language is considered closely intertwined with human cognition, with linguistic structures posited to offer profound insights into the cognitive system (1–3). However, as a coding system, natural language encodes diverse objects into unified forms; its prominent formal features capture people’s attention, such as lexical combinatorial rules, which tend to overshadow those form-independent structures. Here, I present knowledge-level, logic-level, task-level, and model-level semantic structures inherent in natural language. These structures are discovered by shifting the research focus from coding forms of natural language to the objects they encode, unveiling different semantic layers integrated within sentences. The cognitive functions of these structures are evident both in themselves and in models developed from them. I therefore introduce four models to demonstrate their capabilities in memorization, reasoning, learning, natural language generation, and understanding. These findings advance our understanding of natural language and provide a framework for investigating the cognitive system’s information processing through structural analysis of natural language.

1. Introduction

Natural language is a mode of communication, serving as a medium for encoding and transmitting information among individuals (4–6). This perspective raises two fundamental inquiries in studying natural language: What is encoded in natural language (i.e., coding content or meaning)? How is the content encoded (i.e., coding form)? The same content can be encoded with different forms, which means that the distinct arbitrary signs, lexicons, syntaxes, and grammars adopted in different languages actually function as coding forms. Therefore, investigations into linguistic structures grounded in these coding forms primarily address the latter inquiry.

Efforts to explore the former inquiry posit the existence of a semantic system that governs natural language. Various approaches have been developed to delineate and characterize this system, including semantic theories (7–10), semantic web (11), corpus-based semantic models (12–16), and language models (17–19). Despite these approaches’ diligent efforts to mitigate the influence of coding forms on extracting and modeling coding contents, the reality is that they are still more or less conceived, constructed, or trained based on the coding forms. Consequently, limited progress has been achieved in discerning semantic constituents and structures of natural language.

Jerry Fodor argued that the meaning of lexical concepts (i.e., coding content) is determined by their relationships with the

world rather than their relationships with other lexical concepts (20). Inspired by this view, I have reexamined categories of linguistic units (see Tables 1 and 2, Fig. 1E) based on the objects they encode, surpassing the influence of superficial coding forms to unveil the actual coding contents. From these categories, I propose knowledge-level, logic-level, task-level, and model-level semantic structures that structurally decompose natural language according to its coded contents. To further illustrate semantic structure’s roles in shaping and influencing human cognition, I introduce four models grounded in these structures, demonstrating their contributions to memorization, reasoning, learning, natural language generation, and understanding.

- **Mental spatial model (MSM)**, a semantic model (Fig. 2B), reveals how humans encode and recall the spatial position information of entities in the physical world. By offering a relational positioning system that mirrors human spatial cognition, the MSM enables human-machine interaction.

- **Hierarchical coding model (HCM)**, another semantic model (Fig. 3A), uncovers how humans construct conceptual system. It presents a coding system in which various sensory inputs are labeled as elementary attributes that combine to encodes entities in the physical world. Surprisingly, the combination coding strategy employed by the HCM for establishing hierarchical structures explains how humans can effectively learn from small sample sizes.

- **Continuous data model**. By abstracting the general structural features of semantic models like HCM and MSM, I propose a novel type of data mode – the continuous data model. This model integrates a conceptual system with an algebraic system, demonstrating how computation can be applied to reasoning problems and thereby expanding the range of questions that computers may solve.

- **Knowledge base-driven language model (KLM)**. Since sentences are decomposed into four semantic layers by semantic structures, which, together with the establishment of semantic models, inspire a novel framework for addressing natural language processing – the KLM (Fig. 4). It adopts the classic framework of the computer system, comprising two main components: the knowledge base and the program. Within this framework, natural language generation is defined as the process of knowledge acquisition and encoding triggered by specific requests, while learning involves extracting knowledge from natural language and incorporating it into the knowledge base. Moreover, KLM validates and instantiates the hypothesis that understanding is a process (21) that involves acquiring and appropriately using knowledge to satisfy specific goals (22). By validating and operationalizing the foundational assumption in cognitive science that information processing in humans resembles that in computers, KLM provides valuable

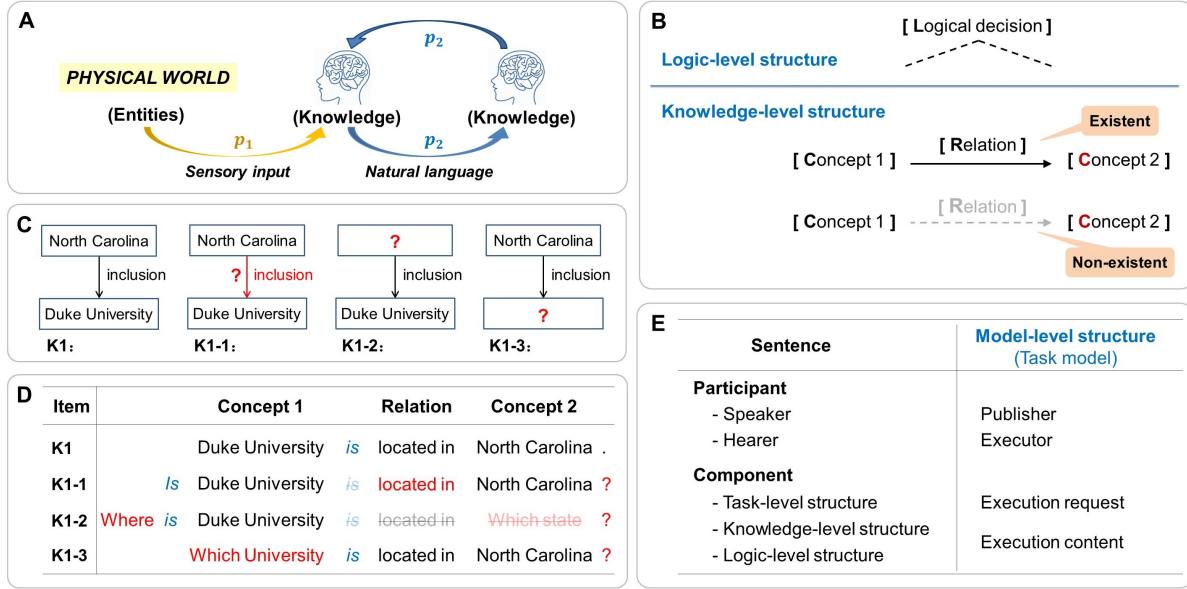


Fig. 1. Schematic representation of the cognitive role and semantic structures of natural language. (A) Diagram illustrating the pathway and carriers involved in information transmission. Sensory organs (eyes, nose, ears, skin, etc.) perceive entities in the physical world through various sensory inputs (information carriers like visible light, sonic waves, and chemical molecules). These inputs are converted into biochemical and bioelectrical signals (the information carriers in human brains) within the brain, which are then processed cognitively to generate knowledge and consciousness. In communication, a portion of this knowledge and consciousness is encoded into natural language (another form of information carrier) to facilitate information exchange among individuals. (B) Schematic of knowledge-level and logic-level structures, highlighting their hierarchical relations. (C) The knowledge segment K1 and its three cases. (D) Examples of task-level structures (English only). These sentences showcase how the data processing requests (highlighted in red in C) are linguistically encoded using iconic words, punctuation, and specific compositional rules. (E) The model-level structure (task model) formalizes tasks encoded by sentences by defining the functions of sentence components.

insights for investigating and interpreting human cognitive activities.

Through these models, I demonstrate the existence of the semantic system underpinning natural language and detail the methods for constructing them. Furthermore, they offer valuable insights into how specific cognitive functions are instantiated within these models, shedding light on the underlying mechanisms of human cognition. This work represents a pioneering formulation of the intricate interplay between natural language, the semantic system, and human cognition.

2. Semantic structures

The relations among natural language, the physical world, human cognition, and knowledge can be depicted as shown in Fig. 1A: Individuals continuously receive sensory inputs from the physical world (process p_1 in Fig. 1A), which undergo cognitive processing to form structured information or data (i.e., knowledge) stored within the brain. These processes occur across diverse environments, resulting in variations in knowledge among individuals. To bridge these knowledge gaps, natural language is employed as the primary tool for encoding and transmitting knowledge between individuals (process p_2 in Fig. 1A). This observation implies that knowledge, a product of human cognition, constitutes one of the contents encoded within natural language.

Knowledge-level structure. Knowledge involves how humans understand and describe the physical world, and it is also defined as a belief that is both true and justified (23, 24). Understanding interrelations among entities in the physical

world pertains to knowledge, which can be depicted by the structure: [Concept 1] + [Relation] + [Concept 2], known as the knowledge-level structure (see Fig. 1C and D). Within this structure, conceptual components represent objects associated with entities, while relational components illustrate relations among these objects.

To identify linguistic units representing these components, I reexamine classifications of linguistic units and propose object-oriented categories (Table. 1), which exemplify three primary subcategories of linguistic units: conceptual, relational, and functional classes. The conceptual class contains words and phrases used to label entities, with the goal of representing the entities themselves and their attributes, which are further categorized into entity words and phrases, as well as various types of attributes (see Fig. 3A), including elementary attributes, extended attributes, advanced attributes, and attribute domains. The relational class contains words and phrases depicting relations among entities. Linguistic units in the functional class represent specific functions and instructions; for example, the word “the” intends to point out specific objects in the context, and the punctuation “?” indicates specific tasks. These object-oriented categories highlight the conceptual and relational components within sentences (see examples in SI Appendix, Fig. S2B), facilitating the extraction of knowledge segments encoded within natural language. Besides, unlike certain coding systems that maintain a strict one-to-one correspondence between coding characters and coding objects (e.g., ASCII and Unicode), natural language often allows a single unit to encode multiple objects, thereby leading to multiple word senses and ambiguity phenomenon.

Logic-level structure. The logic-level structure (Fig. 1B) serves as an additional layer of abstraction above the knowledge-level structure, describing two states of the relation in a knowledge-level structure: existent and non-existent. Words and phrases such as “is/is not”, “have/ have not”, and “can/cannot” function as logical decisions to encode these two states. The usage of the logic-level structure enables the description of the non-existent state of a relation in a knowledge-level structure, thereby doubling the range of knowledge-level structures that natural language can describe (see examples in *SI Appendix, Fig. S2B*). Without the ability to describe the non-existent state of a relation, communication would face significant challenges.

Task-level structure. During communication, individuals alternate between the roles of speaker and hearer. Speakers generate sentences while hearers understand them. The motivation for speakers to produce a sentence may align with the goal of hearers in understanding it. Therefore, I reclassify sentences by examining the speaker’s motivations and objectives for producing them and propose the task-oriented categorization, as detailed in Table 2. The term “task” refers to the implied requests of speakers encoded in sentences, such as knowledge sharing, verification, retrieval, and instruction execution. As exemplified in Fig. 1D, the task-level structure involves how these tasks are encoded in sentences through iconic words, punctuation marks, and specific combinatorial rules. The discovery of task-level structure unveils the essence of language - sentences can be viewed as tasks published by speakers.

Model-level structure. Based on the above findings, I reclassify sentence components by examining their roles in the said tasks and propose a function-oriented categorization of sentence components, as shown in Fig. 1E. This categorization, which defines the functions of sentence components, formalizes the structural framework of tasks, known as the model-level structure. Distinguished from the other three structures, the model-level structure incorporates not only explicit task components expressed in languages, such as execution requests and contents, but also task participants responsible for task publication and execution, which often exist as contextual information. Serving as a task model, the model-level structure suggests a fundamental cognitive framework that formalizes tasks designed by speakers to fulfill specific requirements and assigns them to hearers for execution. This mechanism facilitates communication and cooperation among individuals, thereby fostering community advancement.

3. Two semantic models

Semantic models, or knowledge models, are products of human cognition and can be viewed as mental representations of the physical world. Given that natural language encodes and transmits knowledge stored in human brains, we can reconstruct these semantic models by integrating knowledge segments extracted from natural language. Below, I introduce two models, their construction methods, and associated cognitive functionalities.

MSM. The MSM is constructed based on the knowledge-level structure: [Concept 1] + [Spatial relation] + [Concept 2],

which is extracted from sentences like “Duke University is located in North Carolina”. This structure reveals a spatial positioning strategy humans employ, where individuals use entities with known spatial information to locate target entities. In this structure, “Concept 2” represents the entity that assists in locating “Concept 1”, the target entity. “Spatial relation” refers to the spatial relation between the target and assisting entities. Entities with a relatively stable spatial position (immovable entities) are typically chosen as assisting entities rather than those with constantly changing spatial positions (movable entities). These immovable entities and the spatial relations among them form a stable system that the MSM (Fig. 2B) aims to depict.

Table 3 presents three commonly employed types of spatial relations in natural language, accompanied by their respective reference frames (RF) (25). These include 1) scope relation, encompassing both inclusion and exclusion relations, with the inclusion relation forming $RF_s = \{\overrightarrow{inclusion}\}$; 2) directional relation, further subdivided into 2.1) absolute directional relation consisting of four fixed directions forming $RF_a = \{\overrightarrow{east}, \overrightarrow{west}, \overrightarrow{north}, \overrightarrow{south}\}$, and 2.2) relative directional relation comprising six fixed directions forming $RF_r = \{\overrightarrow{top}, \overrightarrow{bottom}, \overrightarrow{left}, \overrightarrow{right}, \overrightarrow{front}, \overrightarrow{back}\}$; and 3) distance relation. In contrast to RF (26, 27) that rely on “conceptual anchors”, these spatial relation-centered RF exhibit enhanced integrity and universal applicability.

Figure 2B illustrates the MSM architecture, which is structured based on spatial relations, wherein physical entities in the real world are represented as nodes, and their spatial relations are described by directed edges. The MSM integrates graphs (Fig. 2C and D) for depicting spatial directional relations among entities and the tree structure (Fig. 2E) for illustrating spatial scope relations. As a relational positioning system that mirrors human spatial cognition, the MSM can function as a bridge for information exchange between humans and machines (*SI Appendix, Fig. S1*), supplementing existing numerical positioning systems such as GPS. Moreover, it offers excellent flexibility in expanding its structure to accommodate numerous entities, the nodes and edges within the MSM can be easily updated to reflect changes in entities’ spatial positions, while datasets can also be constructed to record trajectories of movable entities. In summary, the MSM empowers machines with memory patterns akin to those observed in humans, thereby facilitating human-machine interactions.

HCM. The HCM presents a coding system that encodes entities in the physical world with attributes perceived by human sensory organs, uncovering how humans construct conceptual systems (Fig. 3A). It is established by examining interrelations among the categories of linguistic units within the conceptual class. In the HCM, nodes represent concepts associated with entities, and directed edges describe inclusion relations among them. Therefore, each node in the HCM is characterized by all its descendant nodes, while a group of child nodes with the same parent node can be represented by this parent node. A node positioned at a higher hierarchical level possesses a greater degree of abstraction, that is, richer semantics. In the HCM, the elementary attribute layer (highlighted in red dotted boxes in Fig. 3A) functions as a transitional or interface layer between the physical world and human memories, where various sensory inputs are labeled as elementary attributes.

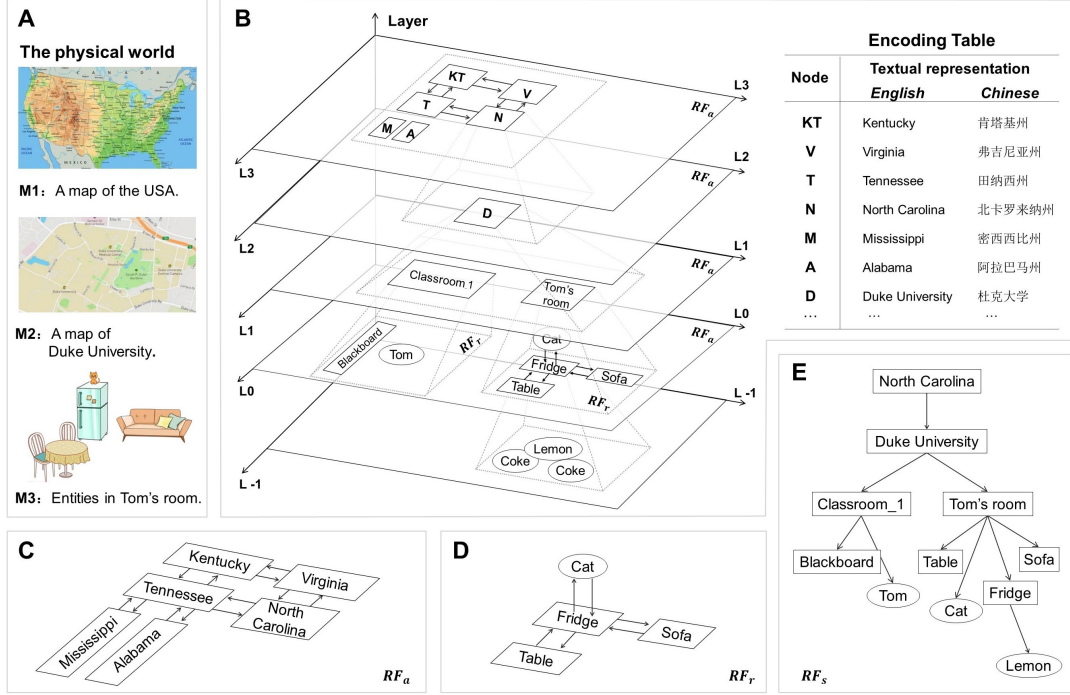


Fig. 2. MSM. (A) The physical world that B to E aims to depict. (B) Schematic of MSM: The vertical structure (inter-layer) illustrates spatial scope relations among entities, while the horizontal structure (intra-layer) depicts spatial directional relations. (C) Graph depicting absolute spatial directional relations among entities shown in A (M1). (D) Graph showing relative spatial directional relations among entities depicted in A (M3). (E) Tree structure illustrating spatial scope relations among entities as presented in A.

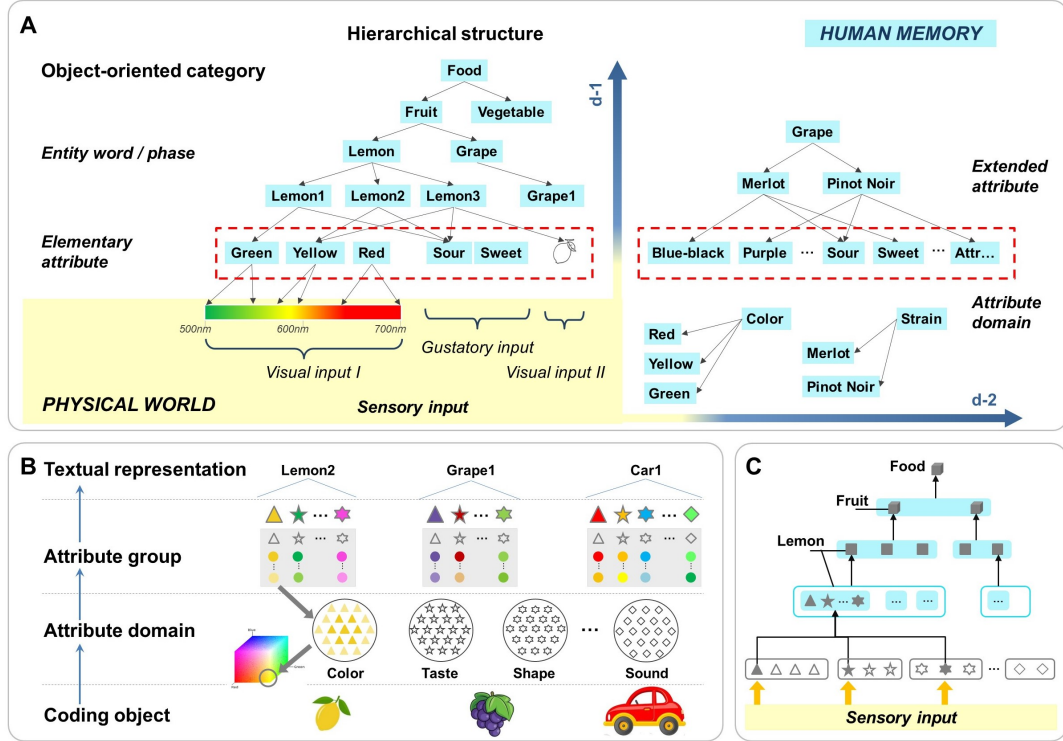


Fig. 3. HCM and Combination Coding. (A) Schematic representation of the hierarchical structure of HCM, illustrating the conceptual inclusion relations among concepts associated with entities. The HCM presents a coding system where various sensory inputs are labeled as elementary attributes that combine to encode entities in the physical world. The d-1 axis denotes the combinational rule based on the physical or conceptual scope of entities, while the d-2 axis indicates the combinational rule based on element categories. (B) Schematic representation of the combination coding strategy, which employs attribute groups (combinations) perceived from physical world entities to encode these entities. (C) Schematic illustrating the application of combination coding in constructing the hierarchical structure of the HCM.

The remaining layers of HCM are structured by employing the combination coding strategy (Fig. 3B and C) introduced below.

Combination coding refers to establishing combinations by selecting $m(m \leq N)$ elements from n sets based on certain rules (each set contains a limited number of distinct elements, and the total number of elements in n sets is N), and utilizing these combinations for encoding specific objects. For instance, the three attribute groups (combinations) shown in Fig. 3B are used to encode the objects “lemon2”, “grape1”, and “car1”. The combination coding suggests that elementary attributes perceived by human sensory systems are employed as elements for encoding physical world entities.

Moreover, the combination coding strategy provides strong generalization performance for the HCM. It is evident in the limited impact that certain range variations of elements within subsets (child nodes) have on the overall combination (parent node). This mechanism is particularly pronounced when the elements within subsets follow a normal distribution. Therefore, combination coding provides theoretical support for humans’ ability to learn from small samples (concept learning), suggesting that a pattern (i.e., a combination) established based on a small sample can represent the majority of the same kind of samples. Additionally, the pattern’s generalization capacity is enhanced with an increasing number of elements within subsets. Given that the human brain, with over ten billion neurons, employs both combination coding and population coding (28) – an established coding strategy in neuroscience – this framework explains humans can efficiently and robustly encode (or memorize) a vast number of physical world entities.

4. Continuous data model

A data model can be classified as a continuous data model if its structure, which refers to the set of relations establishing connections between data elements within the model, is an algebraic system. For instance, the relation set of the HCM: $R_{HCM} = \{\supset\}$, along with the corresponding operation rule $f_{HCM} : \supset + \supset = \supset$, form the algebraic system: (R_{HCM}, f_{HCM}) ; thus, the HCM is a continuous data model.

A continuous data model integrates a conceptual system (i.e., the set of data elements) with an algebraic system (i.e., the set of relations). This means that relations between data elements (i.e., concepts) are computable, enabling the reasoning of unknown knowledge from known knowledge. For example, by leveraging the known knowledge: “*Food $\supset$ Fruit*” and “*Fruit $\supset$ Lemon*”, as shown in Fig. 3A, we can deduce the new knowledge: “*Food $\supset$ Lemon*”.

Structural comparison. The structural comparison of the continuous data model, knowledge graph (29), and widely used relational data model (30) provide insight into their respective advantages in establishing world models. Both the continuous data model and knowledge graph are grounded in graph structures; however, they differ in focus. The continuous data model is relation-centric, distinguished by the same types and computable relations that support relation-based reasoning and search operations. In contrast, the knowledge graph is concept-centric, where data elements are typically linked by diverse types of relations that are not computable,

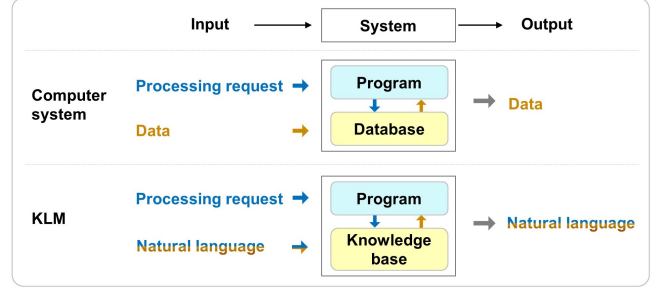


Fig. 4. Comparison of the overall frameworks of computer systems and the KLM. This framework decouples processing objects from their corresponding processing requests, a design widely adopted across various domains, including mechanical and algebraic systems. As the processing object of the KLM, natural language distinguishes itself by encoding both knowledge and its processing requests, necessitating a preprocessing step to separate them.

thereby rendering reasoning operations unfeasible while still ensuring search functionality.

The relational data model employs a table structure, where the term “relation” refers to a two-dimensional table denoted as “*Relation(domain₁, domain₂, ..., domain_n)*”. In contrast to the continuous data model and knowledge graph, which depict stable relations between data elements (concepts), the relational data model focuses on recording changeable attributes of a set of entities. Each row in the table represents a set of changeable attributes of a specific entity, while columns correspond to different attribute types (i.e., attribute domains). These attribute domains, as well as the data elements within them, are mutually independent. Consequently, performing relation-based reasoning operations on a relational data model becomes infeasible. On the positive side, this data independence effectively protects data during numerous accesses and operations, addressing the significant challenge of managing and utilizing large shared databases (30).

5. KLM and its applications

The KLM consists of two main components: the knowledge base and the program (Fig. 4). It adheres to the classical framework employed by computer systems, which separates data from its processing requests (31, 32). The knowledge base functions similarly to human memory, consisting of diverse knowledge models that reconstruct the mental representations of the physical world as perceived and understood by humans, and it is both editable and infinitely expandable. The program component contains instructions designed to manipulate knowledge to achieve specific goals. In contrast to neural network-based language models, the KLM is distinguished by its separate definition and modeling of processing objects (i.e., execution contents) and their corresponding processing requests (i.e., execution requests). This design ensures transparency and traceability, thereby enhancing the model’s credibility and applicability in various scenarios. Subsequent sections will detail the methods for natural language generation, understanding, and learning based on the existing knowledge base comprising MSM and HCM.

Natural language generation. Sentences are tasks published by speakers, which include fixed components such as execution

contents, execution requests, and executors. In the KLM, natural language generation is defined as a sequence of operations aimed at obtaining the task components from corresponding models followed by encoding them into sentences after necessary processing (*SI Appendix, Fig. S2A and B*).

The KLM’s sentence generation logic fundamentally differs from prevailing language models. The latter approach language generation as a statistical inference process based on prior knowledge (i.e., corpora) (33–35), while sentences generated by the KLM encode tasks designed to fulfill specific requirements for higher-order tasks. Therefore, in the KLM, what is encoded in sentences is determined by the requirements rather than their probability of occurrence in prior knowledge. This opens up a new direction for research in natural language generation.

Furthermore, knowledge can be seen as humanity’s collective understanding of the physical world, with knowledge and its representation forms being mutually independent. The KLM framework, which decouples knowledge from its representation forms, enables it to provide translation and multilingual generation services using distinct coding forms. This capability addresses concerns that language models, which are consistently based on majority languages for cost considerations, might exacerbate existing inequalities by disadvantaging speakers of less prevalent languages (36).

Natural language understanding. Understanding an object involves acquiring knowledge about its features, composition, and functions (22, 37). In the KLM, the first two types of knowledge can be obtained through identification operations within relevant knowledge models. For example, by identifying the knowledge “*Sour* $\xleftarrow{\text{inclusion}}$ *Lemon*” and “*Lemon* $\xleftarrow{\text{inclusion}}$ *Fruit*” in the HCM, one can discern the attribute “Sour” and category “Fruit” of the “Lemon”. Similarly, identifying “*Fridge* $\xrightarrow{\text{inclusion}}$ *Lemon*” in the MSM provides location information of the “Lemon”. Object functions, which relate to the object’s role in specific tasks, are defined by relevant task models (Fig. 1E). Consequently, functional knowledge can also be acquired through identification operations within corresponding task models.

Natural language encodes multiple types of objects, including knowledge and tasks. In the KLM, natural language understanding involves initially identifying these objects from sentences and then acquiring relevant knowledge about them (see *SI Appendix, Fig. S2C*). Consequently, understanding is defined as a process consisting of a sequence of identification operations. Within this process, not all identification attempts succeed; the knowledge obtained from successful operations represents the outcomes of understanding. The abundance of these outcomes can be used to gauge the degree of understanding achieved. The extent of knowledge acquired during these understanding processes depends on the number of models available and the volume of relevant data incorporated into them. Therefore, when evaluating KLM’s capability for understanding, the abundance of its knowledge base serves as a more direct and effective indicator than traditional performance metrics (38–40).

Moreover, the further processing of outcomes in understanding constitutes other cognitive activities. For example, the subsequent identification operations performed on the acquired knowledge can be referred to as association (*SI Appendix, Fig. S2C*), while additional manipulation of acquired knowledge,

such as sorting, comparison, and reasoning operations, constitutes advanced cognitive activities like analysis, decision-making, and planning.

Learning. Humans employ diverse learning approaches, such as experiential learning, where knowledge is gained through direct practice (e.g., tasting a lemon to discern its sourness), and indirect learning, where knowledge is obtained from others (e.g., being told that lemons are sour). The KLM exemplifies the latter approach, where learning involves extracting knowledge encoded in natural language and updating the knowledge base accordingly. This process encompasses incorporating (memorizing) new knowledge (*SI Appendix, Fig. S2C*) and revising existing knowledge. Unlike prevailing machine learning methods (41–44), which predominantly extract statistical patterns from massive datasets (45), the KLM introduces a novel learning paradigm distinguished by its cumulative nature and cost-effectiveness, broadening the spectrum of available learning methods.

6. Conclusion

I have proposed four semantic structures, induced by the novel categories of linguistic units, that decompose natural language based on the objects it encodes. Combined with the introduced models and their associated cognitive functions, the feasibility of investigating the human cognitive system through linguistic analysis is demonstrated. I hope further efforts will continue exploring the intersection of natural language, human cognition, and human-like intelligence.

1. MB Everaert, MA Huybregts, N Chomsky, RC Berwick, JJ Bolhuis, Structures, not strings: Linguistics as part of the cognitive sciences. *Trends cognitive sciences* **19**, 729–743 (2015).
2. N Chomsky, *The Minimalist Program*. (MIT Press), (1995).
3. MA Nowak, NL Komarova, P Niyogi, Evolution of universal grammar. *Science* **291**, 114–118 (2001).
4. E Fedorenko, ST Piantadosi, EAF Gibson, Language is primarily a tool for communication rather than thought. *Nature* **630**, 575–586 (2024).
5. E Gibon, et al., How efficiency shapes human language. *Trends Cogn. Sci.* **23**, 389–407 (2019).
6. E Spelke, *Language in Mind: Advances in the Investigation of Language and Thought* eds. D Gentner, S Goldin-Meadow. (MIT Press), pp. 277–311 (2003).
7. S Lappin, C Fox, eds., *The Handbook of Contemporary Semantic Theory*. (Blackwell), (1996).
8. D Geeraerts, *Theories of Lexical Semantics*. (Oxford Univ. Press), (2010).
9. C Goddard, A Wierzbicka, *Words and Meanings: Lexical Semantics Across Domains, Languages, and Cultures*. (Oxford Univ. Press), (2014a).
10. DW Harris, Semantics without semantic content. *Mind & Lang.* **37**, 304–328 (2022).
11. C Fellbaum, ed., *WordNet: an Electronic Lexical Database*. (MIT Press), (1998).
12. K Lund, C Burgess, Producing high-dimensional semantic spaces for lexical co-occurrence. *Behav. Res. Methods* **28**, 203–208 (1996).
13. TK Landauer, ST Dumais, A solution to plato’s problem: The lateen semantic analysis theory of acquisition, induction and representation of knowledge. *Psychol. Rev.* **104**, 211–240 (1997).
14. M Jones, D Mewhort, Representing word meaning and other information in a composite holographic lexicon. *Psychol. Rev.* **114**, 1–37 (2007).
15. K Erk, Vector space models of word meaning and phrase meaning: a survey. *Lang. Linguist. Compas.* **6**, 635–653 (2012).
16. B Lyu, et al., Neural dynamics of semantic composition. *Proc. Natl. Acad. Sci.* **116**, 21318–21327 (2019).
17. J Devlin, MW Chang, K Lee, K Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding in *Proceeding of the 2019 Conference of the North American Chapter of Association for Computational Linguistics: Human Language Technologies*. pp. 4171–4186 (2019).
18. TB Brown, et al., Language models are few-shot learners in *Advances in Neural Information Processing Systems*. pp. 1877–1901 (2020).
19. OpenAI, Gpt-4 technical report, (OpenAI), Technical report (2023).
20. JA Fodor, *LOT2: The Language of Thought Revisited*. (Oxford Univ. Press), (2008).
21. LM Blaha, et al., Understanding is a process. *Front. Syst. Neurosci.* **16**, 1–18 (2022).
22. AR Hough, KA Gluck, The understanding problem in cognitive science. *Adv. Cogn. Syst.* **8**, 13–32 (2019).
23. T Williamson, *Knowledge and its Limits*. (Oxford Univ. Press.), (2002).
24. DP Hunt, The concept of knowledge and how to measure it. *J. Intellect. Cap.* **4**, 100–113 (2003).
25. I Rock, The frame of reference in *The legacy of Solomon asch*. (Psychology Press), pp. 243–268 (2014).

26. E Danziger, Deixis, gesture, and cognition in spatial frame of reference typology. *Stud. Lang.* **34**, 167–185 (2010).
27. H Diessel, S Monakhov, Acquisition of demonstratives in cross-linguistic perspective. *J. Child Lang.* **50**, 922–953 (2023).
28. A Pouget, P Dayan, R Zemel, Information processing with population codes. *Nat. Rev. Neurosci.* **1**, 125–132 (2000).
29. A Hogan, et al., Knowledge graphs. *ACM Comput. Surv.* **54**, 1–31 (2021).
30. EF Codd, A co-relational model of data for large shared data banks. *Commun. ACM* **13**, 377–387 (1970).
31. AM Turing, On computable numbers, with and application to the entscheidungsproblem. *P. Lond. Math. Soc.* **2**, 230–265 (1937).
32. J von Neumann, First draft of a report on the edvac. *IEEE Ann. Hist. Comput.* **2**, 27–75 (1933).
33. A Bender, E. M. Koller, Climbing towards nlu: on meaning, form, and understanding in the age of data in *Proc. 58th Annual Meeting of the Association for Computational Linguistics*. pp. 5185–5198 (2020).
34. ZW Ji, et al., Survey of hallucination in natural language generation. *ACM Comput. Surv.* **55**, 1–38 (2023).
35. R Futrell, Information-theoretic principles in incremental language production. *Proc. Natl. Acad. Sci.* **120**, e2220593120 (2023).
36. A Backus, et al., Minds: big questions for linguistics in the age of ai. *Linguist. Neth.* **40**, 301–308 (2023).
37. HA Simon, Artificial intelligence systems that understand in *International Joint Conference on Artificial Intelligence*. pp. 1059–1073 (1977).
38. P Langley, JE Laird, S Rogers, Cognitive architectures: research issues and challenges. *Cogn. Syst. Res.* **10**, 141–160 (2009).
39. DN Perkins, What is understanding? *Teach. for understanding: Link. research with practice* (1998).
40. HA Simon, *Information-processing Explanations of Understanding*. (Lawrence Erlbaum), (1980).
41. GE Hinton, RR Salakhutdinov, Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
42. Y LeCun, Y. Bengio, G Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
43. V Mnih, et al., Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
44. MI Jordan, TM Mitchell, Machine learning: trends, perspectives, and prospects. *Science* **349**, 255–260 (2015).
45. L Breiman, Statistical modeling: the two cultures. *Stat. Sci.* **16**, 199–231 (2001).
46. J Jara-Ettinger, P Rubio-Fernandez, Demonstratives as attention tools: Evidence of mentalistic representations within language. *Proc. Natl. Acad. Sci.* **121** (2024).

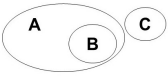
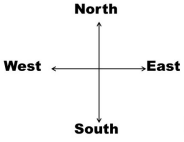
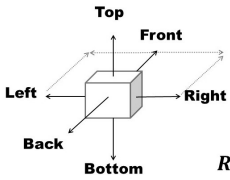
Table 1. Examples of object-oriented categories of linguistic units

Linguistic unit	Category		Coding object
	<i>Conventional</i>	<i>Object-oriented</i>	
The lemon Lemon Yellow, sour	Noun phrase	Entity phrase	The conceptual class - represents the specific entity “lemon” in a specific scenario. - represents the lemon group. Yellow - a label for a segment of visible light perceived by human eyes, which encompasses a wavelength range from 577 to 597 nanometers. Sour – a label for the specific group of molecules and ions perceived by human noses and tongues. Merlot, Pinot Noir - the strains of grapes, which label the collections of selected attributes used to distinguish these two types of grapes, as exemplified in Fig. 3A. Color, strain - represent the categories of attributes. Ripe, unripe - represent the classification results of entities’ ripeness, labeling collections of selected attributes that serve as indicators for ripeness. Squeeze (Type 1 object) - represents the attribute sequences that record the attribute changes of an entity when it is squeezed, e.g., the changes in the contour and pulp of a lemon when it is squeezed. (Type 2 object) - represents the attribute sequences recording the changes in attributes of the agent performing the squeezing action, such as the posture changes of the hand and fingers when executing a lemon squeeze.
	Noun	Entity word	
	Adjective	Elementary attribute	
Merlot, Pinot Noir	Noun	Extended attribute	The functional class (Type 3 object) - the instruction for implementing the squeezing action. It can refer to either a sequence of nervous impulses that coordinate muscle groups to execute a squeezing action or a pre-programmed set of instructions for machines to perform the squeezing action. The, this “.” “?” - symbols that point out specific objects in the context. - segment sentences and indicate their task types .
Color, strain	Noun	Attribute domain	
Ripe, unripe	Adjective	Advanced attribute	
Squeeze	Verb (vt.)	Advanced attribute	The relational class Have, has Of, ‘s Before, after And Because
		Advanced attribute	
		Advanced attribute	
Squeeze	Verb (vt.)	Instruction	The relational class Have, has Of, ‘s Before, after And Because
The, this “.” “?”	Article; pronoun Punctuation	Demonstrative (46) Separator; Task identifier	
Have, has Of, ‘s Before, after And Because	Verb Preposition Preposition Conjunction Conjunction	Inclusion relation Inclusion relation Temporal relation Parallel relation Causal relation	

Table 2. Task-oriented categories of sentences

Category		Coding object
Conventional	Task-oriented	
Declarative & exclamatory sentences	Knowledge sharing sentence	This category of the sentence implies the speaker's expectation that the hearer can memorize the contents shared within sentences. For example, teachers expect students to remember what was taught in class; authors expect readers to understand and retain the concepts and methods presented in their papers. This expectation can be seen as requests(tasks) published by speakers and assigned to hearers for subsequent execution. Compared to declarative sentences, exclamatory sentences typically omit the subject, as it is already implied in the context.
Yes-no question	Knowledge verification sentence	Sentences of this category are generated when speakers intend to verify the existence of a relation within a piece of knowledge, as exemplified in Fig. 1C and D (K1-1). They convey speakers' requests for the hearers to provide answers to these questions.
Wh-question	Knowledge retrieval sentence	Such sentences are formed when speakers lack specific data, as exemplified in Fig. 1C (K1-2 and K1-3). Speakers can employ wh-words to replace the missing data and adjust the sentence structure accordingly (refer to K1-2 and K1-3 in Fig. 1D) to convey their requests for hearers to provide the requested data.
Imperative sentence	Instruction sentence	Instruction sentences are typically structured without explicitly stating the executors (i.e., subjects); instead, the intended hearers are implicitly designated as the executors to carry out the instructions represented by verbs. Note that the verbs that convey these instructions correspond to the <i>Type 3 object</i>, as previously detailed in Table 1.

Table 3. The classification of spatial relations employed in natural language

Spatial relations	Lexical representations	Reference system
1.Scope relations	Inclusion: in, at... Exclusion: outside of...	 RF_s
2.1 Absolute directional relations	East: east of... West: west of... North: the north side of ... South: the south side of...	 RF_a
2.2 Relative directional relations	Top: on, above, over, on top of... Bottom: below, under, beneath... Left: left of... Right: the right side of... Front: before, in front of... Back: behind, back of...	 RF_r
3.Distance relations	by, near, next to, beside...	